# Convergence of Algorithms of Decomposition Type for the Eigenvalue Problem

D. S. Watkins*†
*Department of Pure and Applied Mathematics*
*Washington State University*
*Pullman, Washington 99164-2930*

and

L. Elsner
*Fakultät für Mathematik*
*Universität Bielefeld*
*Postfach 86 40*
*D-4800 Bielefeld 1, Federal Republic of Germany*

Submitted by Richard A. Brualdi

ABSTRACT

We develop the theory of convergence of a generic *GR* algorithm for the matrix eigenvalue problem that includes the *QR*, *LR*, *SR*, and other algorithms as special cases. Our formulation allows for shifts of origin and multiple *GR* steps. The convergence theory is based on the idea that the *GR* algorithm performs nested subspace iteration with a change of coordinate system at each step. Thus the convergence of the *GR* algorithm depends on the convergence of certain sequences of subspaces. It also depends on the quality of the coordinate transformation matrices, as measured by their condition numbers. We show that with a certain obvious shifting strategy the *GR* algorithm typically has a quadratic asymptotic convergence rate. For matrices possessing certain special types of structure, cubic convergence can be achieved.

## 1. INTRODUCTION

The *QR* and *LR* algorithms are well-known procedures for calculating eigenvalues and eigenvectors of matrices. There are other not so well-known

*Supported by the National Science Foundation under grant DMS-8800437.
†watkins@wsumath.bitnet.



 0024-3795/91/$3.50

algorithms, e.g. *SR* and *HR*, which can be useful in special situations. All of these algorithms have very similar theories, and they also have similar practical implementations. The results are scattered in the literature. We felt it would be useful to develop a general theory that includes all of these algorithms as special cases. A step in this direction was taken by Della-Dora [14], who proved a convergence theorem for algorithms based on subgroup decompositions. His theorem covered the *QR*, *LR*, and other algorithms simultaneously, but it was limited to the unshifted case and dealt only with matrices whose eigenvalues have distinct moduli. Our objective is to study the algorithms as they are really used, i.e. with varying shifts and multiple steps. In this paper we present a general convergence theory. In a subsequent article [25] we will discuss issues associated with the implementation of implicit versions of the algorithms.

We begin by introducing (in Section 2) our object of study, a generic *GR* algorithm, and listing several examples. The *GR* algorithm is an iterative procedure that begins with a matrix $A$, whose eigenvalues we would like to know, and produces a sequence of similar matrices $(A_i)$ that (hopefully) converges to upper triangular form, exposing the eigenvalues. The transforming matrices for the similarity transformations $A_i = G_i^{-1} A_{i-1} G_i$ are obtained from a "*GR* decomposition" $p_i(A_{i-1}) = G_i R_i$, in which $R_i$ is upper triangular and $p_i$ is a polynomial. The degree of $p_i$ is called the *multiplicity* of the $i$th step. Until recently workers have focused their attention on single and double steps, i.e. steps of multiplicity one and two, respectively. In recent years it has been recognized that steps of higher multiplicity are sometimes useful. For example, in the *SR* algorithm [9] it is natural to use steps of multiplicity four. Recently Bai and Demmel [1] have experimented with *QR* steps of multiplicity as high as 20, the objective being to improve the opportunities for parallelism in a *QR* step. Since we allow steps of any multiplicity, our theory covers all of these cases.

In Section 3 we show that every *GR* algorithm is a form of nested subspace iteration in which a change of coordinate system is made at each step. This insight is the key to a clear understanding of why the algorithms converge. The connection between subspace iteration and the *LR* algorithm has been known for a long time. It was pointed out by Bauer in his earliest work [3] on *Treppeniteration*, a form of subspace iteration. The connection between the *QR* algorithm and subspace iteration was reported in Wilkinson's book [26] and elsewhere, but its utility as a vehicle for understanding the *QR* and similar algorithms and proving that they converge seems not to have been appreciated until Buurema's dissertation [11]. The message was subsequently reinforced by Parlett and Poole [21] and again by Watkins [24].

In Section 4 we present several simple lemmas concerning distances between subspaces. These are used in Section 5 to help prove our basic

convergence theorems for subspace iteration. While proofs of convergence of subspace iteration have been given before (e.g. [11, 21]), they have focused on the unshifted case. Our theorems are concerned with nonstationary, i.e. variable-shift, subspace iteration. These are then used in Section 6 to prove theorems about the convergence of the *GR* algorithm. It is interesting that the convergence of all *GR* algorithms is based on the convergence of the same sequences of subspaces, assuming the same sequence of shifts is chosen in each case.[1] What sets the various algorithms apart from one another is the varying quality of the transforming matrices $G_i$ used to perform the change of coordinates at each step. Our theorems guarantee convergence only if the condition numbers of the accumulated transforming matrices $\hat{G}_i = G_1 G_2 \cdots G_i$ remain bounded as the iterations proceed.

Our global convergence theorem holds for shifting strategies that converge. Unfortunately no one has ever been able to devise a practical shifting strategy that is guaranteed to converge for all matrices and can be shown to converge rapidly.[2] Indeed, there appears to be little hope for a universally valid, global convergence theorem for shifting strategies that are used in practice. Batterson and Smillie [2] have even shown that one well-known strategy, the Rayleigh-quotient shift, can behave chaotically. The set of matrices on which chaotic behavior occurs has positive Lebesgue measure in the space $\mathbb{C}^{n\times n}$. It may well be that other shift strategies can also exhibit chaotic behavior, although this is not something that has been observed frequently.

For local convergence the situation is better. We are able to show that for a particular practical strategy, which we call the generalized Rayleigh-quotient strategy, the local convergence rate is typically quadratic. For matrices having certain types of special structure, it is cubic. For example, the *QR* algorithm applied to a normal matrix typically converges cubically. This is a known result, at least for the case of single and double-step algorithms. As a second example, the *SR* algorithm applied to Hamiltonian matrices typically converges cubically.

Earlier proofs of the quadratic convergence of the *QR* algorithm have been based on the fact that the *QR* algorithm can be viewed as inverse iteration, in particular Rayleigh-quotient iteration. Our approach makes no reference to inverse iteration whatsoever. That such an approach should be possible is made clear by the duality theorem discussed in [24, Theorem 4.1],

[1] Similar observations have been made previously by Bauer [4] and Parlett and Poole [21].

[2] Success has been achieved for the important special case of tridiagonal Hermitian matrices. On this class, the *QR* algorithm with the Wilkinson shift strategy has rapid, global convergence. See [20, §8-10].

which illustrates the fundamental connection between direct and inverse subspace iteration.

## 2. THE GENERIC *GR* ALGORITHM

Our results will be stated in terms of a generic *GR* algorithm, which is in turn based on a generic *GR* decomposition. A *GR decomposition* is any well-defined rule by which every matrix $C$ in some large class of matrices $\mathscr{C}$ can be expressed as a product

$$C = GR,$$

where $G$ is nonsingular and $R$ is upper triangular. In other words, a *GR* decomposition is a rule that assigns to each $C \in \mathscr{C}$ a unique nonsingular $G$ that "reduces $C$ to triangular form," in the sense that $G^{-1}C$ is upper triangular. This second statement of the definition reflects the way in which *GR* decompositions are often computed in practice. Given a *GR* decomposition, we can define a corresponding *GR* algorithm, in fact, a whole family of algorithms. Let $A$ be a matrix whose eigenvalues we would like to know. The *GR* algorithm generates a sequence $(A_i)$ of similar matrices as follows. $A_0$ is taken to be $A$ or some convenient matrix similar to $A$, say $A_0 = G_0^{-1}AG_0$. Given $A_{i-1}$, let $p_i$ be some polynomial such that $p_i(A_{i-1}) \in \mathscr{C}$. Then $p_i(A_{i-1})$ has a *GR* decomposition: $p_i(A_{i-1}) = G_iR_i$. Define $A_i$ by $A_i = G_i^{-1}A_{i-1}G_i$. The step from $A_{i-1}$ to $A_i$ can be expressed succinctly by the two equations

$$p_i(A_{i-1}) = G_iR_i, \tag{1}$$

$$A_i = G_i^{-1}A_{i-1}G_i. \tag{2}$$

Under suitable conditions the sequence $(A_i)$ will tend to upper triangular, or at least block triangular, form, yielding information about the eigenvalues. Information about eigenvectors and invariant subspaces is obtained by accumulating the transforming matrices $G_i$. The choice of the $p_i$ has much to do with the rate of convergence. The simplest choice is $p_i(A_{i-1}) = A_{i-1}$, which yields the *basic GR* algorithm. If we wish to have the sequence converge rapidly, we must make a cleverer choice. The *Rayleigh-quotient* shifting strategy often works well. We take $p_i(A_{i-1}) = A_{i-1} - \sigma_i I$, where the *shift* $\sigma_i$ is taken to be the $(n, n)$ entry of $A_{i-1}$. Another good choice is $p_i(A_{i-1}) = (A_{i-1} - \sigma_i)(A_{i-1} - \tau_i)$, where $\sigma_i$ and $\tau_i$ are the eigenvalues of the lower

right-hand $2\times 2$ submatrix of $A_{i-1}$. Both of these strategies are special cases of the generalized Rayleigh-quotient strategy to be discussed in Section 6.

The degree of $p_i$ is called the *multiplicity* of the $i$th step. If $p_i$ has degree 1, it is a *single* step. If the degree is 2, it is a *double* step, and so on. Writing $p_i$ in factored form: $p_i(A) = \alpha_i(A-\sigma_1^{(i)})(A-\sigma_2^{(i)})\cdots(A-\sigma_m^{(i)})$, we call the roots $\sigma_1^{(i)},\ldots,\sigma_m^{(i)}$ the *shifts* for the $i$th step. Each step of multiplicity $m$ has $m$ shifts. A procedure for choosing the $p_i$ is called a *shifting strategy* because the choice of $p_i$ implies a certain choice of shifts $\sigma_1^{(i)},\ldots,\sigma_m^{(i)}$. The $p_i$ are usually chosen to be monic ($\alpha_i = 1$). This is a minor point, for all of the $GR$ algorithms that we will consider have the following property: If $p(A)$ has the decomposition $p(A) = GR$, then for any $\alpha \neq 0$ the $GR$ decomposition of $\alpha p(A)$ is $G\tilde{R}$, where $\tilde{R} = \alpha R$. This property implies that the outcome of a $GR$ step is invariant under rescaling of $p$.

If $p_i(A_{i-1})$ is nonsingular, then $R_i$ must be nonsingular, and it is easily shown that $A_i = R_i A_{i-1} R_i^{-1}$. Therefore, if $A_{i-1}$ is in upper Hessenberg form, $A_i$ will also be in upper Hessenberg form. It is common to choose $G_0$ so that $A_0$ is in upper Hessenberg form. Then all $A_i$ will be in upper Hessenberg form, as long as all $p_i(A_{i-1})$ are nonsingular. While nonsingularity is the rule, singular $p_i(A_{i-1})$ do not cause problems in practice; in fact they are good news.

EXAMPLE 2.1 ($QR$ decomposition). Let $\mathscr{E} = \mathbb{C}^{n\times n}$. Every $C \in \mathscr{E}$ can be expressed as a product $C = QR$, where $Q$ is unitary and $R$ is upper triangular. One can specify rules for calculating $Q$ and $R$ so that they are uniquely determined. For example, one could say that $A$ is to be reduced to upper triangular form by reflectors, as in Algorithm 5.2.1 of [17]. The $QR$ decomposition gives rise to the famous $QR$ algorithm.

EXAMPLE 2.2 ($LR$ decomposition). Let $\mathscr{E} \subseteq \mathbb{C}^{n\times n}$ be the set of matrices whose $n-1$ leading principal minors are nonzero. Every $C \in \mathscr{E}$ can be expressed uniquely as a product $C = LR$, where $L$ is unit lower triangular and $R$ is upper triangular. This decomposition and that of the following example give rise to variants of the $LR$ algorithm.

EXAMPLE 2.3 ($LR$ decomposition with partial pivoting). Let $\mathscr{E} = \mathbb{C}^{n\times n}$. Every $C \in \mathscr{E}$ can be expressed as a product $C = KR$, where $K$ and $R$ are uniquely determined by the rules of Gaussian elimination with partial pivoting, as determined by e.g. the subroutine SGEFA from LINPACK [15]. The matrix $R$ is upper triangular, and $K$ has the form $K = P_1 L_1 P_2 L_2 \cdots P_{n-1} L_{n-1}$, where each $L_i$ is a Gauss transformation whose entries all have

modulus less than or equal to 1, and each $P_i$ is either a transposition or the identity matrix.

EXAMPLE 2.4 ($SR$ decomposition). Define $J \in \mathbb{R}^{2n\times 2n}$ by $J = \operatorname{diag}\{\tilde{J}, \tilde{J}, \ldots, \tilde{J}\}$, where

$$\tilde{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

A matrix $S \in \mathbb{R}^{2n\times 2n}$ is *symplectic* if $S^T J S = J$. (This is the *shuffled* form of symplectic matrices.) Let $\mathscr{C}$ be the set of $C \in \mathbb{R}^{2n\times 2n}$ such that the leading principal minors of $C^T J C$ of even order are all nonzero. Every $C \in \mathscr{C}$ can be expressed as a product $C = SR$, where $S$ is symplectic and $R$ is upper triangular [16, Theorem 11; 7, Satz 4.5.11]. Algorithms for producing unique factors $S$ and $R$ are given in [7], [9], and [10]. The $SR$ decomposition gives rise to the $SR$ algorithm, which can be used to solve the algebraic Riccati equation [9, 10].

EXAMPLE 2.5 ($HR$ decomposition). Let $J \in \mathbb{C}^{n\times n}$ be a diagonal matrix with $\pm 1$'s on the main diagonal, and let $\mathscr{C} \in \mathbb{C}^{n\times n}$ be the set of all matrices with nonzero leading principal minors. Every $C \in \mathscr{C}$ can be expressed as a product $C = HR$, where $R$ is upper triangular, and $H$ satisfies $H^* J H = P^T J P$ for some permutation $P$ [6, 16, 8, 7]. The $HR$ decomposition gives rise to the $HR$ algorithm, a generalization of the $QR$ algorithm that can be applied effectively to certain eigenvalue problems.

EXAMPLE 2.6 (Complex orthogonal $QR$ decomposition). In certain applications [12] one needs the eigenvalues of complex, symmetric (*not* Hermitian) matrices. For this purpose a complex orthogonal $QR$ decomposition is useful. Almost all $A \in \mathbb{C}^{n\times n}$ can be expressed as a product $A = QR$, where $R$ is upper triangular and $Q$ is complex and orthogonal (*not* unitary). The resulting complex orthogonal $QR$ algorithm preserves the complex symmetry property. See Cullum and Willoughby [12] for details.

## 3. THE $GR$ ALGORITHM AS SUBSPACE ITERATION

The most important thing to understand about the $GR$ algorithm is that it is a form of nested subspace iteration in which a change of coordinate system is made at each step. In the basic version of subspace iteration, we choose a

subspace $\mathcal{S}$ and form a sequence of subspaces $(\mathcal{S}_i)$ by $\mathcal{S}_0 = \mathcal{S}$ and

$$\mathcal{S}_i = A\mathcal{S}_{i-1}, \qquad i = 1,2,3,\ldots.$$

Then $\mathcal{S}_i = A^i\mathcal{S}$ for all $i$. This is just a multidimensional form of the basic power method. Under mild conditions on $A$ and $\mathcal{S}$ the $\mathcal{S}_i$ converge to an invariant subspace of $A$. In an effort to improve the convergence rate, one can consider a nonstationary subspace iteration scheme in which $A$ is replaced by a shifted matrix $A - \sigma_i I$ at the $i$th step. We will consider even more general nonstationary schemes of the form

$$\mathcal{S}_i = p_i(A)\mathcal{S}_{i-1}, \qquad i = 1,2,3,\ldots,$$

in which $(p_i)$ is some sequence of polynomials. Letting $\hat{p}_i = p_i \cdots p_2 p_1$, we have

$$\mathcal{S}_i = \hat{p}_i(A)\mathcal{S}, \qquad i = 1,2,3,\ldots. \tag{3}$$

In Section 5 we will state and prove some precise conditions under which $(\mathcal{S}_i)$ converges to an invariant subspace of $A$.

Now let's see how the $GR$ algorithm can be interpreted as subspace iteration. The $i$th step of the $GR$ algorithm begins with the $GR$ decomposition

$$p_i(A_{i-1}) = G_i R_i. \tag{4}$$

To keep the discussion uncomplicated, we will assume $p_i(A_{i-1})$ is nonsingular. Let $g_1, g_2, \ldots, g_n$ denote the columns of $G_i$, and let $e_1, e_2, \ldots, e_n$ be the standard basis vectors in $\mathbb{C}^n$. Since $R_i$ is upper triangular and nonsingular, it follows easily from (4) that for every $k \in \{1,2,\ldots,n\}$, the space spanned by the first $k$ columns of $p_i(A_{i-1})$ is the same as the space spanned by the first $k$ columns of $G_i$; that is,

$$p_i(A_{i-1})\langle e_1,\ldots,e_k\rangle = \langle g_1,\ldots,g_k\rangle.$$

Thus $\langle g_1,\ldots,g_k\rangle$ is the space obtained from one step of subspace iteration, starting from $\langle e_1,\ldots,e_k\rangle$. To finish the $GR$ step we perform the similarity transformation

$$A_i = G_i^{-1} A_{i-1} G_i,$$

which we view as a change of coordinate system. $A_{i-1}$ and $A_i$ are representatives of the same linear transformation with respect to two different

coordinate systems. Given a coordinate vector $x$ that is the representation of some "physical" vector with respect to the old $A_{i-1}$ system, the coordinate vector of the same physical vector with respect to the new $A_i$ coordinate system is $G_i^{-1}x$. Thus the vectors that are represented by $g_1,\ldots,g_k$ in the old system are given by $G_i^{-1}g_1,\ldots,G_i^{-1}g_k$ in the new system. But the latter are obviously just $e_1,\ldots,e_k$. Thus the space that is represented by $\langle g_1,\ldots,g_k\rangle$ in the old coordinate system is represented by $\langle e_1,\ldots,e_k\rangle$ in the new system. To summarize: Each step of the *GR* algorithm performs one step of subspace iteration on $\langle e_1,\ldots,e_k\rangle$, resulting in some space $\langle g_1,\ldots,g_k\rangle$. A change of coordinate system then maps $\langle g_1,\ldots,g_k\rangle$ back to $\langle e_1,\ldots,e_k\rangle$. After this transformation we are ready for the next step, which performs another step of subspace iteration on this same space. We conclude that the *GR* algorithm performs a sequence of steps of subspace iteration, starting with $\mathcal{S}_0=\langle e_1,\ldots,e_k\rangle$. Furthermore, at each step it performs a change of coordinates, so that the space $\mathcal{S}_i$ obtained after $i$ steps is represented by $\langle e_1,\ldots,e_k\rangle$ with respect to the newest coordinate system. Thus, instead of having a fixed matrix and a sequence of subspaces, we have a fixed subspace and a sequence of matrices. If the subspace iterations converge as hoped, $\langle e_1,\ldots,e_k\rangle$ will become closer and closer to being an invariant subspace of $A_i$. If it were exactly an invariant subspace of some $A_i$, that $A_i$ would have the block triangular form

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \tag{5}$$

where $A_{11}\in\mathbb{C}^{k\times k}$. But $\langle e_1,\ldots,e_k\rangle$ is not exactly an invariant subspace of any of the $A_i$, so what typically occurs is that $(A_i)$ approaches the form (5) as $i\to\infty$.

So far we haven't specified $k$. In fact the subspace iteration takes place for all $k\in\{1,2,\ldots,n\}$ simultaneously. Thus we actually have a nested family of subspace iterations. If each of these subspace iterations converges, then $(A_i)$ will (typically) tend to the form (5) for all values of $k$ simultaneously; that is, $(A_i)$ will tend to upper triangular form, exposing the eigenvalues on the main diagonal. Even if convergence fails for some values of $k$, the sequence $(A_i)$ will still converge to a block triangular form that will usually be useful. Precise conditions under which convergence can be guaranteed are given in Section 6.

## 4. DISTANCES BETWEEN SUBSPACES

Given a subspace $\mathcal{S}$ of $\mathbb{C}^n$, let $P_{\mathcal{S}}$ denote the orthoprojector of $\mathbb{C}^n$ onto $\mathcal{S}$. To gauge the convergence of sequences of subspaces we will define the

following standard metric [18, 17] on the set of subspaces of $\mathbb{C}^n$:

$$d(\mathscr{S}, \mathscr{T}) = \|P_{\mathscr{S}} - P_{\mathscr{T}}\|_2.$$

Since $P_{\mathscr{S}^\perp} = I - P_{\mathscr{S}}$, we see immediately that $d(\mathscr{S}^\perp, \mathscr{T}^\perp) = d(\mathscr{S}, \mathscr{T})$. An equivalent definition is

$$d(\mathscr{S}, \mathscr{T}) = \sup_{\substack{s \in \mathscr{S} \\ \|s\|_2 = 1}} d(s, \mathscr{T}) = \sup_{\substack{s \in \mathscr{S} \\ \|s\|_2 = 1}} \inf_{t \in \mathscr{T}} \|s - t\|_2 \tag{6}$$

if $\dim(\mathscr{S}) = \dim(\mathscr{T})$, and $d(\mathscr{S}, \mathscr{T}) = 1$ otherwise.

Given a matrix $S \in \mathbb{C}^{n \times k}$, let $R(S)$ denote the range or column space of $S$.

LEMMA 4.1. *Let $\mathscr{S}$ and $\mathscr{T}$ be two $k$-dimensional subspaces of $\mathbb{C}^n$, and let $S \in \mathbb{C}^{n \times k}$ be a matrix with orthonormal columns such that $\mathscr{S} = R(S)$. Then there exists $T \in \mathbb{C}^{n \times k}$ with orthonormal columns such that $\mathscr{T} = R(T)$ and $\|S - T\|_2 \leqslant \sqrt{2}\, d(\mathscr{S}, \mathscr{T})$.*

*Proof.* $\mathscr{S}$ and $\mathscr{T}$ have orthonormal bases of principal vectors [5, 17] $\tilde{s}_1, \ldots, \tilde{s}_k$ and $\tilde{t}_1, \ldots, \tilde{t}_k$, respectively, such that

$$(\tilde{s}_i, \tilde{t}_j) = 0 \qquad \text{if} \quad i \neq j, \tag{7}$$

and the angle between $\tilde{s}_i$ and $\tilde{t}_i$ is $\theta_i$, the $i$th principal angle between $\mathscr{S}$ and $\mathscr{T}$. The principal angles satisfy $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_k \leqslant \pi/2$, and $\sin \theta_k = d(\mathscr{S}, \mathscr{T})$. Therefore, for all $i$,

$$\|\tilde{s}_i - \tilde{t}_i\|_2 = 2 \sin(\theta_i/2) \leqslant \sqrt{2} \sin \theta_i \leqslant \sqrt{2}\, d(\mathscr{S}, \mathscr{T}).$$

Let $\tilde{S} = [\tilde{s}_1, \ldots, \tilde{s}_k]$ and $\tilde{T} = [\tilde{t}_1, \ldots, \tilde{t}_k] \in \mathbb{C}^{n \times k}$. Using (7), we see that

$$\|\tilde{S} - \tilde{T}\|_2 = \max_{1 \leqslant i \leqslant k} \|\tilde{s}_i - \tilde{t}_i\|_2 \leqslant \sqrt{2}\, d(\mathscr{S}, \mathscr{T}).$$

Since $R(S) = R(\tilde{S})$, and both $S$ and $\tilde{S}$ have orthonormal columns, there exists a unitary $U \in \mathbb{C}^{k \times k}$ such that $S = \tilde{S}U$. Let $T = \tilde{T}U$. Then $R(T) = R(\tilde{T}) = \mathscr{T}$, and $\|S - T\|_2 = \|\tilde{S}U - \tilde{T}U\|_2 = \|\tilde{S} - \tilde{T}\|_2 \leqslant \sqrt{2}\, d(\mathscr{S}, \mathscr{T})$. ∎

LEMMA 4.2. *Let $\mathscr{S}$ and $\mathscr{T}$ be two subspaces of $\mathbb{C}^n$ of the same dimension, let $V \in \mathbb{C}^{n \times n}$ be nonsingular, and let $\hat{\mathscr{S}} = V^{-1}\mathscr{S}$ and $\hat{\mathscr{T}} =$*

$V^{-1}\mathcal{T}$. *Then*

$$d(\mathcal{S},\mathcal{T}) \leq \kappa_2(V)\, d(\tilde{\mathcal{S}},\tilde{\mathcal{T}}).$$

*Proof.* In (6) the supremum is attained, so there exists $s \in \mathcal{S}$ with $\|s\|_2 = 1$ such that $d(\mathcal{S},\mathcal{T}) = d(s,\mathcal{T})$. Let $\hat{s} = V^{-1}s$, $\sigma = \|\hat{s}\|_2$, and $\tilde{s} = \sigma^{-1}\hat{s}$, so that $\|\tilde{s}\|_2 = 1$. Notice that $\sigma \leq \|V^{-1}\|_2$. Pick $\tilde{t} \in \tilde{\mathcal{T}}$ such that $\|\tilde{s} - \tilde{t}\|_2 = d(\tilde{s},\tilde{\mathcal{T}})$. (This is an infimum, but it also is attained.) Since $\|\tilde{s}\|_2 = 1$, $d(\tilde{s},\tilde{\mathcal{T}}) \leq d(\tilde{\mathcal{S}},\tilde{\mathcal{T}})$. Let $\hat{t} = \sigma\tilde{t}$ and $t = V\hat{t}$. Then $d(\mathcal{S},\mathcal{T}) = d(s,\mathcal{T}) \leq \|s-t\|_2 = \|V(\sigma(\tilde{s}-\tilde{t}))\|_2 \leq \|V\|_2 \|V^{-1}\|_2 \|\tilde{s}-\tilde{t}\|_2 \leq \kappa_2(V) d(\tilde{\mathcal{S}},\tilde{\mathcal{T}})$. ■

LEMMA 4.3. *Let $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{T}}$ be subspaces of $\mathbb{C}^n$ of the same dimension, and let $\tilde{\mathcal{U}}$ be the orthogonal complement of $\tilde{\mathcal{T}}$. Then $\tilde{\mathcal{S}} \cap \tilde{\mathcal{U}} = \{0\}$ if and only if $d(\tilde{\mathcal{S}},\tilde{\mathcal{T}}) < 1$.*

The proof of Lemma 4.3 is an easy exercise.

The final lemma of this section begins to deal with subspace iteration.

LEMMA 4.4. *Let $\tilde{\mathcal{T}} = \langle e_1,\ldots,e_k\rangle$ and $\tilde{\mathcal{U}} = \langle e_{k+1},\ldots,e_n\rangle$, let $\tilde{\mathcal{S}} \subseteq \mathbb{C}^n$ be a $k$-dimensional space such that $\tilde{\mathcal{S}} \cap \tilde{\mathcal{U}} = \{0\}$, and let $\beta = d(\tilde{\mathcal{S}},\tilde{\mathcal{T}}) < 1$. Let $T \in \mathbb{C}^{n\times n}$ be a block diagonal matrix $T = \operatorname{diag}\{T_1, T_2\}$, where $T_1 \in \mathbb{C}^{k\times k}$; let $p$ be a polynomial such that $p(T_1)$ is nonsingular; and let $\tilde{\mathcal{S}}' = p(T)\tilde{\mathcal{S}}$. Then*

$$d(\tilde{\mathcal{S}}',\tilde{\mathcal{T}}) \leq \frac{\beta}{\sqrt{1-\beta^2}} \|p(T_2)\|_2 \|p(T_1)^{-1}\|_2.$$

*Proof.* Given $x = [x_1, x_2, \ldots, x_n]^T \in \tilde{\mathcal{S}}$, let $x' = [x_1,\ldots,x_k]^T \in \mathbb{C}^k$ and $x'' = [x_{k+1},\ldots,x_n]^T \in \mathbb{C}^{n-k}$. Clearly $d(x,\tilde{\mathcal{T}}) = \|x''\|_2$ and $d(x,\tilde{\mathcal{U}}) = \|x'\|_2$. We will begin by demonstrating that

$$\|x''\|_2 \leq \frac{\beta}{\sqrt{1-\beta^2}} \|x'\|_2 \tag{8}$$

for all $x \in \tilde{\mathcal{S}}$. We can assume, without loss of generality, that $\|x\|_2 = 1$. Since $\|x''\|_2 = d(x,\tilde{\mathcal{T}})$, we have $\|x''\|_2 \leq \beta$. Since $\|x'\|_2^2 + \|x''\|_2^2 = 1$, we also have $\|x'\|_2 \geq \sqrt{1-\beta^2} > 0$. The inequality (8) follows.

Given any $x \in \mathscr{S}$ and $y = p(T)x$, we have

$$y' = p(T_1)x', \qquad y'' = p(T_2)x''.$$

Since $p(T_1)$ is nonsingular, $x' = p(T_1)^{-1}y'$, and

$$\|x'\|_2 \leqslant \|p(T_1)^{-1}\|_2 \|y'\|_2. \tag{9}$$

Now choose $y \in p(T)\mathscr{S} = \mathscr{S}'$ such that $\|y\|_2 = 1$ and $d(\mathscr{S}', \tilde{\mathscr{T}}) = d(y, \tilde{\mathscr{T}})$. Then there exists $x \in \mathscr{S}$ such that $y = p(T)x$, and

$$d(\mathscr{S}', \tilde{\mathscr{T}}) = \|y''\|_2 \leqslant \|p(T_2)\|_2 \|x''\|_2. \tag{10}$$

The proof is completed by combining (10), (8), and (9), and noting that $\|y'\|_2 \leqslant 1$. ∎

## 5. CONVERGENCE OF SUBSPACE ITERATION

We present two theorems on the convergence of nonstationary subspace iterations. The first concerns simple matrices.

THEOREM 5.1. *Let $A \in \mathbb{C}^{n\times n}$ be a simple matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (in any convenient order) and associated linearly independent eigenvectors $v_1, v_2, \ldots, v_n$. Let $V = [v_1\ v_2\ \cdots\ v_n] \in \mathbb{C}^{n\times n}$, and let $\kappa_2(V)$ denote the condition number of $V$ with respect to the spectral norm. Let $k$ be some integer satisfying $1 \leqslant k \leqslant n-1$, and define invariant subspaces $\mathscr{T} = \langle v_1, \ldots, v_k \rangle$ and $\mathscr{U} = \langle v_{k+1}, \ldots, v_n \rangle$. Let $(p_i)$ be a sequence of polynomials, and let $\hat{p}_i = p_i \cdots p_2 p_1$ for all $i$. Suppose*

$$\hat{p}_i(\lambda_j) \neq 0, \qquad j = 1, \ldots, k, \tag{11}$$

*for all $i$, and let*

$$r_i = \frac{\max\limits_{k+1 \leqslant j \leqslant n} |\hat{p}_i(\lambda_j)|}{\min\limits_{1 \leqslant j \leqslant k} |\hat{p}_i(\lambda_j)|}.$$

*Let $\mathcal{S}$ be any $k$-dimensional subspace of $\mathbb{C}^n$ satisfying*

$$\mathcal{S} \cap \mathcal{U} = \{0\}. \tag{12}$$

*Let $\mathcal{S}_i = \hat{p}_i(A)\mathcal{S}$, $i = 1,2,\ldots$, as in (3). Then there exists a constant $C$ (depending on $\mathcal{S}$) such that for all $i$,*

$$d(\mathcal{S}_i, \mathcal{T}) \leqslant C\kappa_2(V)r_i.$$

*In particular $\mathcal{S}_i \to \mathcal{T}$ if $r_i \to 0$.*

REMARK. The subspace condition (12) is almost certain to be satisfied by a subspace $\mathcal{S}$ chosen at random, since $\dim(\mathcal{S}) + \dim(\mathcal{U}) = \dim(\mathbb{C}^n)$. In the context of the *GR* algorithm we will be able to guarantee that (12) is satisfied.

EXAMPLE 5.2. Consider a stationary iteration in which $p_i(A) = p(A)$ for all $i$. [If we take $p(A) = A$, we have basic subspace iteration.] Then $\hat{p}_i(A) = p(A)^i$. Order $\lambda_1,\ldots,\lambda_n$, so that $|p(\lambda_1)| \geqslant |p(\lambda_2)| \geqslant \cdots \geqslant |p(\lambda_n)|$. Suppose $k$ is such that

$$\rho = \frac{\max\limits_{k+1 \leqslant j \leqslant n} |p(\lambda_j)|}{\min\limits_{1 \leqslant j \leqslant k} |p(\lambda_j)|} = \frac{|p(\lambda_{k+1})|}{|p(\lambda_k)|} < 1.$$

Then $r_i = \rho^i$, so $\mathcal{S}_i \to \mathcal{T}$ linearly with the contraction number $\rho$.

EXAMPLE 5.3. Given $k$, suppose we can order $\lambda_1,\ldots,\lambda_n$ so that the sets $\{\lambda_1,\ldots,\lambda_k\}$ and $\{\lambda_{k+1},\ldots,\lambda_n\}$ are disjoint. Let $m = n - k$, and for $j = 1,\ldots,m$ let $(\sigma_j^{(i)})$ be a sequence such that $\sigma_j^{(i)} \to \lambda_{k+j}$ as $i \to \infty$, and $\sigma_j^{(i)} \notin \{\lambda_1,\ldots,\lambda_k\}$ for all $i$. For each $i$ let $p_i$ be the polynomial of degree $m$ given by

$$p_i(\lambda) = (\lambda - \sigma_1^{(i)})(\lambda - \sigma_2^{(i)}) \cdots (\lambda - \sigma_m^{(i)}).$$

Let $p(\lambda) = (\lambda - \lambda_{k+1})(\lambda - \lambda_{k+2}) \cdots (\lambda - \lambda_n)$. Then, as $i \to \infty$,

$$p_i(\lambda_j) \to p(\lambda_j) \quad \begin{cases} = 0 & \text{if} \quad k+1 \leqslant j \leqslant n, \\ \neq 0 & \text{if} \quad 1 \leqslant j \leqslant k. \end{cases}$$

Therefore for every $\rho > 0$ there exists $i_0$ such that for all $i > i_0$,

$$\frac{\max\limits_{k+1 \leq j \leq n} |p_i(\lambda_j)|}{\min\limits_{1 \leq j \leq k} |p_i(\lambda_j)|} \leq \rho.$$

It follows that there is a constant $K$ such that $r_i \leq K\rho^i$ for all $i$. This is true for every $\rho > 0$, so $r_i \to 0$ superlinearly. Therefore $\mathcal{S}_i \to \mathcal{T}$ superlinearly.

*Proof of Theorem 5.1.* The assumption (11) implies that any null vectors that $\hat{p}_i(A)$ might have must lie in $\mathcal{U}$. Thus we see from (12) that $\mathcal{S}$ contains no null vectors of $\hat{p}_i(A)$. It follows that $\mathcal{S}_i = \hat{p}_i(A)\mathcal{S}$ has dimension $k$ for all $i$. Therefore the distances $d(\mathcal{S}_i, \mathcal{T})$ are given by (6). Let $\tilde{\mathcal{S}} = V^{-1}\mathcal{S}$, $\tilde{\mathcal{S}}_i = V^{-1}\mathcal{S}_i$, $\tilde{\mathcal{T}} = V^{-1}\mathcal{T} = \langle e_1, \ldots, e_k \rangle$, $\tilde{\mathcal{U}} = V^{-1}\mathcal{U} = \langle e_{k+1}, \ldots, e_n \rangle$, and $D = V^{-1}AV = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$. Then $\tilde{\mathcal{S}}_i = \hat{p}_i(D)\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}} \cap \tilde{\mathcal{U}} = \{0\}$. By Lemma 4.2,

$$d(\mathcal{S}_i, \mathcal{T}) \leq \kappa_2(V)\, d(\tilde{\mathcal{S}}_i, \tilde{\mathcal{T}}). \tag{13}$$

It now suffices to prove the theorem for the diagonal matrix $D$.

Let $T_1 = \operatorname{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{C}^{k \times k}$ and $T_2 = \operatorname{diag}\{\lambda_{k+1}, \ldots, \lambda_n\} \in \mathbb{C}^{(n-k)\times(n-k)}$, so that $D = \operatorname{diag}\{T_1, T_2\}$. By (11) $\hat{p}_i(T_1)$ is nonsingular, so by Lemma 4.4,

$$d(\tilde{\mathcal{S}}_i, \tilde{\mathcal{T}}) \leq C \|\hat{p}_i(T_2)\|_2 \|\hat{p}_i(T_1)^{-1}\|_2,$$

where $C = d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}}) / \sqrt{1 - d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}})^2}$. Combining this inequality with (13), and noting that $\|\hat{p}_i(T_2)\|_2 = \max_{k+1 \leq j \leq n} |\hat{p}_i(\lambda_j)|$ and $\|\hat{p}_i(T_1)^{-1}\|_2 = [\min_{1 \leq j \leq k} |\hat{p}_i(\lambda_j)|]^{-1}$, we are done. ∎

REMARKS. The size of the constant $C\kappa_2(V)$ is of some practical importance. If $A$ is normal, $V$ can be chosen so that $\kappa_2(V) = 1$. Otherwise $\kappa_2(V) > 1$. As $A$ approaches a defective matrix, $\kappa_2(V) \to \infty$. The constant

$$C = \frac{d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}})}{\sqrt{1 - d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}})^2}}$$

is a measure of how well $\tilde{\mathcal{S}}$ (the transformed initial guess) approximates the invariant (under $D$) subspace $\tilde{\mathcal{T}}$. We can make $C$ arbitrarily close to zero by taking $\tilde{\mathcal{S}}$ sufficiently close to $\tilde{\mathcal{T}}$. On the other hand, $C \to \infty$ as $d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}}) \to 1$.

Since virtually every matrix that arises in practice is simple, Theorem 5.1 is adequate for most needs. However, in order to make our coverage more complete, we will prove another theorem in the same spirit that is valid for both simple and defective matrices. Here our hypotheses are more restrictive, but they are satisfied in typical applications of the *GR* algorithm.

THEOREM 5.4. *Let $A \in \mathbb{C}^{n\times n}$, and let $p$ be a polynomial of degree $\leqslant n$. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of $A$, ordered so that $|p(\lambda_1)| \geqslant |p(\lambda_2)| \geqslant \cdots \geqslant |p(\lambda_n)|$. Suppose $k$ is a positive integer less than $n$ for which $|p(\lambda_k)| > |p(\lambda_{k+1})|$, let $\rho = |p(\lambda_{k+1})|/|p(\lambda_k)|$, and let $(p_i)$ be a sequence of polynomials of degree $\leqslant n$ such that $p_i \to p$ as $i \to \infty$ and $p_i(\lambda_j) \neq 0$ for $j = 1, \ldots, k$ and all $i$. Let $\mathcal{T}$ and $\mathcal{U}$ be the invariant subspaces of $A$ associated with $\lambda_1, \ldots, \lambda_k$ and $\lambda_{k+1}, \ldots, \lambda_n$, respectively. Consider the nonstationary subspace iteration*

$$\mathcal{S}_i = p_i(A)\mathcal{S}_{i-1},$$

*where $\mathcal{S}_0 = \mathcal{S}$ is a $k$-dimensional subspace of $\mathbb{C}^n$ satisfying $\mathcal{S} \cap \mathcal{U} = \{0\}$. Then for every $\hat{\rho}$ satisfying $\rho < \hat{\rho} < 1$ there is a constant $\hat{C}$ such that*

$$d(\mathcal{S}_i, \mathcal{T}) \leqslant \hat{C}\hat{\rho}^i, \qquad i = 1, 2, 3, \ldots.$$

REMARK. By $p_i \to p$ we mean convergence with respect to the unique norm topology on the finite-dimensional space of polynomials of degree $\leqslant n$. This hypothesis implies that $p_i(M) \to p(M)$ for any complex matrix $M$.

REMARK. The theorem shows that convergence is at least linear in the contraction number $\hat{\rho}$. Rapid convergence can be achieved by taking $p$ to have degree $m = n - k$, say $p(x) = (x - \sigma_1)\cdots(x - \sigma_m)$, where $\sigma_1, \ldots, \sigma_m$ are chosen to be good approximations to $\lambda_{k+1}, \ldots, \lambda_n$. The optimal choice is $\sigma_i = \lambda_{k+l}$ for $l = 1, \ldots, m$, since then $\rho = 0$, and the convergence is superlinear.

*Proof.* For all $i$, $\mathcal{S}_i = \hat{p}_i(A)\mathcal{S}$, where $\hat{p}_i = p_i \cdots p_2 p_1$, as before. As in the proof of Theorem 5.1, the assumption that $p_i(\lambda_j) \neq 0$ for $j = 1, \ldots, k$ and the subspace condition $\mathcal{S} \cap \mathcal{U} = \{0\}$ imply together that $\mathcal{S}_i$ has dimension $k$ for all $i$.

Let $V \in \mathbb{C}^{n\times n}$ be any nonsingular matrix such that $V^{-1}AV$ is a block diagonal matrix of the form $T = \operatorname{diag}\{T_1, T_2\}$, where $T_1 \in \mathbb{C}^{k\times k}$ and $T_2 \in \mathbb{C}^{(n-k)\times(n-k)}$ are upper triangular matrices with main-diagonal entries $\lambda_1, \ldots, \lambda_k$ and $\lambda_{k+1}, \ldots, \lambda_n$, respectively. For example, $\operatorname{diag}\{T_1, T_2\}$ could be the Jordan canonical form of $A$, or it could be obtained from the Schur form $\hat{T}$ by a similarity transformation $W^{-1}\hat{T}W$, where $W$ has the form

$$\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}$$

(cf. [17, Lemma 7.1.5]). Then $\mathcal{T} = \langle v_1, \ldots, v_k \rangle$ and $\mathcal{U} = \langle v_{k+1}, \ldots, v_n \rangle$, where $v_1, \ldots, v_n$ are the columns of $V$. Let $\tilde{\mathcal{T}} = V^{-1}\mathcal{T} = \langle e_1, \ldots, e_k \rangle$, $\tilde{\mathcal{U}} = V^{-1}\mathcal{U} = \langle e_{k+1}, \ldots, e_n \rangle$, $\tilde{\mathcal{S}} = V^{-1}\mathcal{S}$, and $\tilde{\mathcal{S}}_i = V^{-1}\mathcal{S}_i$. Then $\tilde{\mathcal{S}} \cap \tilde{\mathcal{U}} = \{0\}$ and $\tilde{\mathcal{S}}_i = \hat{p}_i(T)\tilde{\mathcal{S}}$. The condition $p_i(\lambda_j) \neq 0$ for all $i$ and for $j = 1, \ldots, k$ implies that $\hat{p}_i(T_1)$ is nonsingular. Thus by Lemmas 4.2 and 4.4,

$$d(\mathcal{S}_i, \mathcal{T}) \leq C\kappa_2(V)\|\hat{p}_i(T_2)\|_2\|\hat{p}_i(T_1)^{-1}\|_2, \tag{14}$$

where $C = d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}})/\sqrt{1 - d(\tilde{\mathcal{S}}, \tilde{\mathcal{T}})^2}$. Let $\nu = |p(\lambda_{k+1})|$ and $\delta = |p(\lambda_k)|$, so that $\rho = \nu/\delta$. Given $\hat{\rho}$ with $\rho < \hat{\rho} < 1$, choose $\bar{\rho}$ so that $\rho < \bar{\rho} < \hat{\rho}$. There is a unique $\epsilon$ such that $0 < \epsilon < \delta$ and $\bar{\rho} = (\nu + \epsilon)/(\delta - \epsilon)$. Let $\bar{\nu} = \nu + \epsilon$ and $\bar{\delta} = \delta - \epsilon > 0$, so that $\bar{\rho} = \bar{\nu}/\bar{\delta}$. There exists $i_0$ such that for all $i > i_0$, $\max_{k+1 \leq j \leq n} |p_i(\lambda_j)| \leq \bar{\nu}$ and $\min_{1 \leq j \leq k} |p_i(\lambda_j)| \geq \bar{\delta}$. Let $C_2 = \|\hat{p}_{i_0}(T_2)\|_2$ and $C_1 = \|\hat{p}_{i_0}(T_1)^{-1}\|_2$. Then

$$\|\hat{p}_i(T_2)\|_2 \leq C_2 \left\| \prod_{j=i_0+1}^{i} p_j(T_2) \right\|_2$$

and

$$\|\hat{p}_i(T_1)^{-1}\|_2 \leq C_1 \left\| \prod_{j=i_0+1}^{i} p_j(T_1)^{-1} \right\|_2.$$

The matrix $p(T_2)$ is upper triangular, so it can be expressed as a sum $p(T_2) = D + N$, where $D = \operatorname{diag}\{p(\lambda_{k+1}), \ldots, p(\lambda_n)\}$ and $N$ is strictly upper triangular. Each of the matrices $p_j(T_2)$ has an analogous representation $p_j(T_2) = D_j + N_j$. Since $p_j \to p$ (in any norm) as $j \to \infty$, we have also $p_j(T_2)$

$\to p(T_2)$, $D_j \to D$, and $N_j \to N$. Thus there is a constant $K$ such that $\|N_j\|_2 \leq K$ for all $j$. Also, by definition of $i_0$, $\|D_j\|_2 \leq \bar{\nu}$ for all $j > i_0$. Now

$$\prod_{j=i_0+1}^{i} p_j(T_2) = \prod_{j=i_0+1}^{i} (D_j + N_j),$$

which can be rewritten as a sum of $2^{i-i_0}$ terms, each of which is a product of $i - i_0$ diagonal ($D_j$) and strictly upper triangular ($N_j$) matrices. Since the matrices are of order $m$, each term that has $m$ or more strictly upper triangular factors must be zero. For each $h < m$, each term having $h$ strictly upper triangular factors must also have $i - i_0 - h$ diagonal factors. The norm of each of the strictly upper triangular factors is bounded above by $K$, the diagonal factors by $\bar{\nu}$, so the norm of the entire term is bounded above by $K^h \bar{\nu}^{i-i_0-h}$. For each $h < m$ there are $\binom{i-i_0}{h}$ terms having exactly $h$ strictly upper triangular factors, so

$$\|\hat{p}_i(T_2)\|_2 \leq C_2 \sum_{h=0}^{m-1} \binom{i-i_0}{h} K^h \bar{\nu}^{i-i_0-h}$$

$$= \bar{\nu}^i C_2 \sum_{h=0}^{m-1} \binom{i-i_0}{h} K^h \bar{\nu}^{-i_0-h}.$$

Since $\binom{i-i_0}{h}$ is a polynomial in $i$ of degree $h$, the second sum is just a polynomial. Thus

$$\|\hat{p}_i(T_2)\|_2 \leq \pi_2(i)\bar{\nu}^i, \tag{15}$$

where $\pi_2$ is a polynomial of degree at most $m-1$. The same sort of argument can be applied to $\|\hat{p}_i(T_1)^{-1}\|_2$. The matrix $p(T_1)^{-1}$ is upper triangular and can be expressed as a sum $p(T_1)^{-1} = E + M$, where $E = \operatorname{diag}\{p(\lambda_1)^{-1}, \ldots, p(\lambda_k)^{-1}\}$ and $M$ is strictly upper triangular. Each $p_j(T_1)^{-1}$ has an analogous representation $p_j(T_1)^{-1} = E_j + M_j$. Since $p_j(T_1) \to p(T_1)$ as $j \to \infty$, we also have $p_j(T_1)^{-1} \to p(T_1)^{-1}$, $E_j \to E$, and $M_j \to M$. Thus there exists $K'$ such that $\|M_j\|_2 \leq K'$ for all $j$. Furthermore, for $j > i_0$, $\|E_j\|_2 \leq \bar{\delta}^{-1}$. Thus, reasoning as before,

$$\|\hat{p}_i(T_1)^{-1}\|_2 \leq \pi_1(i)\bar{\delta}^{-i}, \tag{16}$$

where $\pi_1$ is a polynomial in $i$ of degree at most $k-1$. Combining (14), (15), and (16) we find that

$$d(\mathcal{S}_i, \mathcal{T}) \leq C\kappa_2(V)\pi(i)\bar{\rho}^i,$$

where $\pi$ is a polynomial. Now $\pi(i)\bar{\rho}^i = [\pi(i)(\bar{\rho}/\hat{\rho})^i]\hat{\rho}^i$. Since $0 < \bar{\rho}/\hat{\rho} < 1$, and $\pi$ is a mere polynomial, $\pi(i)(\bar{\rho}/\hat{\rho})^i \to 0$ as $i \to \infty$. In particular this factor is bounded. Therefore there exists $\hat{C}$ such that $d(\mathcal{S}_i, \mathcal{T}) \leq \hat{C}\hat{\rho}^i$ for all $i$. ■

## 6. CONVERGENCE OF THE *GR* ALGORITHM

Consider the *GR* algorithm (1,2), starting from $A_0$. For $i = 1,2,3,\ldots$ define the accumulated transforming matrices

$$\hat{G}_i = G_1 G_2 \cdots G_i, \qquad \hat{R}_i = R_i \cdots R_2 R_1.$$

Then from (2),

$$A_i = \hat{G}_i^{-1} A_0 \hat{G}_i \tag{17}$$

for all $i$. The accumulated transforming matrices also satisfy the fundamental identity

$$\hat{p}_i(A_0) = \hat{G}_i \hat{R}_i, \tag{18}$$

where $\hat{p}_i = p_i \cdots p_2 p_1$, as before. This is easily proved by induction. For $i = 1$ it coincides with the case $i = 1$ of (1). For $i = j > 1$ assume (18) holds for $i = j-1$. Then $\hat{p}_j(A_0) = p_j(A_0)\hat{G}_{j-1}\hat{R}_{j-1} = \hat{G}_{j-1}[\hat{G}_{j-1}^{-1}p_j(A_0)\hat{G}_{j-1}]\hat{R}_{j-1}$ $= \hat{G}_{j-1}p_j(A_{j-1})\hat{R}_{j-1} = \hat{G}_{j-1}G_jR_j\hat{R}_{j-1} = \hat{G}_j\hat{R}_j$. Equation (18) is the key to our analysis. It not really anything new; every convergence proof of which we are aware for algorithms of this type makes use of an equation like (18). However, our use of (18) differs from the way it has been used previously. In the *QR* case, (18) gives the unique *QR* decomposition of $\hat{p}_i(A_0)$. Wilkinson's convergence proof in [26], which is typical, is based upon this fact and the continuity of the decomposition $C = QR$ as a function of (nonsingular) $C$. The same approach can be applied to the *LR* algorithm without pivoting, the *SR* algorithm, and various other *GR* algorithms whose *GR* decomposition is defined by membership of the factor $G$ in some closed subgroup of $GL_n(\mathbb{C})$

(e.g. unitary, unit lower triangular, or symplectic group). See Della-Dora [14]. However, there are certain important algorithms to which this approach does not apply, for example the *LR* algorithm with partial pivoting. In our approach, by contrast, the hypotheses do not imply that (18) gives the *GR* decomposition of $\hat{p}_i(A_0)$, nor do they imply that the *GR* decomposition is continuous. Since we do not rely on these properties, our analysis applies to the *LR* algorithm with pivoting, as well as the other algorithms.

In Section 3 we observed, by looking at the algorithm one step at a time, that the *GR* algorithm is just subspace iteration. Equation (18) yields the same observation in cumulative form. Assuming (for simplicity) that $\hat{p}_i(A_0)$ is nonsingular, (18) shows that the space spanned by the first $k$ columns of $\hat{G}_i$ is just $\hat{p}_i(A_0)\langle e_1,\ldots,e_k\rangle$. Being the result of $i$ steps of subspace iteration starting from $\langle e_1,\ldots,e_k\rangle$, this space is (hopefully) close to an invariant subspace, in which case $A_i = \hat{G}_i^{-1}A_0\hat{G}_i$ is (hopefully) close to block triangular form. Notice that the sequence of subspaces $\hat{p}_i(A_0)\langle e_1,\ldots,e_k\rangle$ is determined by the choice of polynomials $p_i$ and does not depend explicitly on which *GR* decomposition (e.g. *QR*, *LR*, *SR*, etc.) is being used. The choice of *GR* decomposition affects the sequence of matrices $A_i$ through the cumulative transformation matrices $\hat{G}_i$. We cannot guarantee convergence unless these are reasonably well behaved. Precise conditions for convergence are given in Theorem 6.2, which relies heavily upon the following lemma.

LEMMA 6.1. *Let $A \in \mathbb{C}^{n\times n}$, and let $\mathcal{T} \subseteq \mathbb{C}^n$ be a $k$-dimensional space that is invariant under $A$. Let $G \in \mathbb{C}^{n\times n}$ be a nonsingular matrix, and let $\mathcal{S}$ be the space spanned by the first $k$ columns of $G$. (Think of $\mathcal{S}$ as an approximation to $\mathcal{T}$.) Let $B = G^{-1}AG$, and consider the partitioned form*

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

*where $B_{21} \in \mathbb{C}^{(n-k)\times k}$. Then*

$$\|B_{21}\|_2 \leq 2\sqrt{2}\,\kappa_2(G)\|A\|_2 d(\mathcal{S},\mathcal{T}).$$

*Proof.* Consider the decomposition $G = QR$, where $Q$ is unitary and $R$ is upper triangular. Making the partition $Q = [Q_1\ Q_2]$, where $Q_1 \in \mathbb{C}^{n\times k}$, we have $\mathcal{S} = R(Q_1)$ and $\mathcal{S}^\perp = R(Q_2)$. Using the partition

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, \quad \text{where} \quad R_{11} \in \mathbb{C}^{k\times k},$$

one easily verifies that $B_{21} = R_{22}^{-1} Q_2^* A Q_1 R_{11}$. Noting that $\|R_{11}\|_2 \leq \|R\|_2 = \|G\|_2$ and $\|R_{22}^{-1}\|_2 \leq \|R^{-1}\|_2 = \|G^{-1}\|_2$, we conclude that

$$\|B_{21}\|_2 \leq \kappa_2(G) \|Q_2^* A Q_1\|_2. \tag{19}$$

By Lemma 4.1 there exist $T_1 \in \mathbb{C}^{n \times k}$ and $T_2 \in \mathbb{C}^{n \times m}$ with orthonormal columns, such that $R(T_1) = \mathscr{T}$ and $R(T_2) = \mathscr{T}^{\perp}$,

$$\|Q_1 - T_1\|_2 \leq \sqrt{2}\, d(\mathscr{S}, \mathscr{T}), \quad \text{and} \quad \|Q_2 - T_2\|_2 \leq \sqrt{2}\, d(\mathscr{S}^{\perp}, \mathscr{T}^{\perp}).$$

Now $Q_2^* A Q_1$ can be rewritten as

$$Q_2^* A Q_1 = (Q_2 - T_2)^* A Q_1 + T_2^* A (Q_1 - T_1) + T_2^* A T_1.$$

Since $\mathscr{T}$ is invariant under $A$, $T_2^* A T_1 = 0$. Therefore

$$\begin{aligned} \|Q_2^* A Q_1\|_2 &\leq \|Q_2 - T_2\|_2 \|A\|_2 \|Q_1\|_2 + \|T_2\|_2 \|A\|_2 \|Q_1 - T_1\|_2 \\ &\leq 2\sqrt{2}\, \|A\|_2 d(\mathscr{S}, \mathscr{T}). \end{aligned} \tag{20}$$

Combining (19) and (20), we are done. ∎

THEOREM 6.2. *Let $A_0 \in \mathbb{C}^{n \times n}$, and let $p$ be a polynomial. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of $A_0$, ordered so that $|p(\lambda_1)| \geq |p(\lambda_2)| \geq \cdots \geq |p(\lambda_n)|$. Suppose $k$ is a positive integer less than $n$ such that $|p(\lambda_k)| > |p(\lambda_{k+1})|$, let $\rho = |p(\lambda_{k+1})| / |p(\lambda_k)|$, and let $(p_i)$ be a sequence of polynomials such that $p_i \to p$ and $p_i(\lambda_j) \neq 0$ for $j = 1, \ldots, k$ and all $i$. Let $\mathscr{T}$ and $\mathscr{U}$ be the invariant subspaces of $A_0$ associated with $\lambda_1, \ldots, \lambda_k$ and $\lambda_{k+1}, \ldots, \lambda_n$, respectively, and suppose $\langle e_1, \ldots, e_k \rangle \cap \mathscr{U} = \{0\}$. Let $(A_i)$ be the sequence of iterates of the GR algorithm using these $p_i$, starting from $A_0$. If there exists a constant $\hat{\kappa}$ such that the cumulative transformation matrices $\hat{G}_i$ all satisfy $\kappa_2(\hat{G}_i) \leq \hat{\kappa}$, then $(A_i)$ tends to block triangular form, in the following sense. Write*

$$A_i = \begin{bmatrix} A_{11}^{(i)} & A_{12}^{(i)} \\ A_{21}^{(i)} & A_{22}^{(i)} \end{bmatrix},$$

*where $A_{11}^{(i)} \in \mathbb{C}^{k \times k}$. Then for every $\hat{\rho}$ satisfying $\rho < \hat{\rho} < 1$ there exists a constant $C$ such that $\|A_{21}^{(i)}\|_2 \leq C \hat{\rho}^i$ for all $i$.*

REMARK. It follows that the eigenvalues of $A_{11}^{(i)}$ and $A_{22}^{(i)}$ converge to $\lambda_1,\dots,\lambda_k$ and $\lambda_{k+1},\dots,\lambda_n$, respectively, as can be shown by standard techniques.

*Proof.* Let $\mathcal{S} = \langle e_1,\dots,e_k\rangle$ and $\mathcal{S}_i = \hat{p}_i(A_0)\mathcal{S}$ for all $i$. All of the hypotheses of Theorem 5.4 are satisfied, so for every $\hat{\rho}$ satisfying $\rho < \hat{\rho} < 1$ there exists $\hat{C}$ such that $d(\mathcal{S}_i, \mathcal{T}) \leq \hat{C}\hat{\rho}^i$ for all $i$. Consider the partition $\hat{G}_i = [\hat{G}_1^{(i)}\ \hat{G}_2^{(i)}]$, where $\hat{G}_1^{(i)} \in \mathbb{C}^{n\times k}$. As we remarked above, (18) implies that the columns of $\hat{G}_1^{(i)}$ span $\mathcal{S}_i$; that is, $\mathcal{S}_i = R(\hat{G}_1^{(i)})$. This is true regardless of whether or not $\hat{p}_i(A_0)$ is nonsingular, since $\mathcal{S}_i$ and $R(\hat{G}_1^{(i)})$ have the same dimension. Applying Lemma 6.1 with the roles of $A$, $G$, and $B$ played by $A_0$, $\hat{G}_i$, and $A_i$, respectively, we conclude that

$$\|A_{21}^{(i)}\|_2 \leq 2\sqrt{2}\,\kappa_2(\hat{G}_i)\|A_0\|_2\hat{C}\hat{\rho}^i \leq C\hat{\rho}^i,$$

where $C = 2\sqrt{2}\,\hat{\kappa}\|A_0\|_2\hat{C}$. ∎

It is usually the case that the hypotheses of the theorem hold for many values of $k$ simultaneously, and the sequence $(A_i)$ converges to a corresponding block triangular form. In the ideal case, in which the hypotheses are satisfied for all $k$, the limiting form is upper triangular.

In practice the $p_i$ are polynomials of some low degree $m \ll n$. When the $GR$ algorithm functions as intended (and this is usually the case), the limiting polynomial $p$ has eigenvalues of $A$ as its roots. Thus, if we take $k = n - m$, then $\rho = 0$, and the iterates will effectively converge after just a few steps to the block triangular form

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

with $A_{22} \in \mathbb{C}^{m\times m}$. Since $m$ is small, it is a simple matter to compute the eigenvalues of $A_{22}$. The rest of the eigenvalues of $A$ are eigenvalues of $A_{11}$, so subsequent $GR$ steps can be applied to the reduced matrix $A_{11}$. Since the hypotheses of Theorem 6.2 are typically satisfied not just for $k = n - m$, but also for many values of $k < n - m$, $A_{11}$ will almost certainly already have made some progress toward convergence. Thus the remaining eigenvalues will be easier to extract.

We noted earlier that the $GR$ algorithm is generally applied to upper Hessenberg matrices. An upper Hessenberg matrix is called *irreducible* if all of its subdiagonal entries are nonzero. Given an upper Hessenberg matrix that is not irreducible, we can reduce its eigenvalue problem to two or more

subproblems involving irreducible matrices, so we can restrict our attention, without loss of generality, to irreducible upper Hessenberg matrices. For this class of matrices the subspace condition $\langle e_1, \ldots, e_k \rangle \cap \mathscr{U} = \{0\}$ is automatically satisfied [19]. Indeed, suppose $x \in \langle e_1, \ldots, e_k \rangle$ is nonzero. Its last nonzero component is $x_r$, where $r \leqslant k$. Since $A_0$ has irreducible upper Hessenberg form, the last nonzero component of $A_0 x$ is its $(r+1)$st, the last nonzero component of $A_0^2 x$ is its $(r+2)$nd, and so on. It follows that $x, A_0 x, A_0^2 x, \ldots, A_0^m x$ are linearly independent, where $m = n - k$. Therefore the smallest invariant subspace of $A_0$ that contains $x$ has dimension at least $m+1$. Since $\mathscr{U}$ is invariant under $A_0$ and has dimension $m$, $x \notin \mathscr{U}$. Thus $\langle e_1, \ldots, e_k \rangle \cap \mathscr{U} = \{0\}$.

Theorem 6.2 shows that there are two questions whose answers determine whether or not the *GR* algorithm converges: (1) Can we choose $p_i$ so that the subspaces converge? (2) Can the condition numbers $\kappa_2(\hat{G}_i)$ be kept under control? Unfortunately, we cannot answer either of these questions definitively, except in special cases. We will discuss the second question first. In the case of the *QR* algorithm the conditioning of the transforming matrices is not a problem; the $\hat{G}_i$ are unitary, so they satisfy $\kappa_2(\hat{G}_i) = 1$. Thus one can concentrate on question (1). In all other cases steps must be taken to control the condition numbers.

In the *LR* algorithm with partial pivoting, the interchanges are a heuristic attempt to control $\kappa_2(\hat{G}_i)$. Each $\hat{G}_i$ has the form $P_1 L_1 P_2 L_2 P_3 L_3 \cdots P_{i(n-1)} L_{i(n-1)}$, where the $P_j$ are permutations, and the $L_j$ are Gauss transformations whose multipliers have modulus no greater than 1. The $P_j$ all have condition number 1, and each $L_j$ is reasonably well conditioned: $\kappa_2(L_j) \leqslant n\kappa_1(L_j) \leqslant 4n$. Unfortunately this does not guarantee that the product of many such transformations will be well conditioned. In practice the condition number usually does remain at a reasonable level, and the algorithm (usually) works quite well.

An exception to this last statement is given by matrices of the form

$$A_{i-1} = C = \begin{bmatrix} a & b \\ b & 0 \end{bmatrix},$$

where $|a| < |b|$; cf. [26, p. 511]. A single step with shift zero, i.e. $p_i(A_{i-1}) = A_{i-1}$, yields $A_i = A_{i-1}$; the algorithm is stationary. This is not due to failure of the underlying subspaces to converge; they usually do converge. For example, if $a$ and $b$ are real and $a \neq 0$, $C$ has real eigenvalues of distinct modulus, so the subspace iterations converge. The failure can be attributed to growth of the condition numbers of the transforming matrices. An easy calculation shows that the transforming matrix $G_i$ is just $b^{-1}C$. If the step is

repeated many times, we have $\hat{G}_i = b^{-i}C^i$, so $\kappa_2(\hat{G}_i) = \kappa_2(C^i) = \kappa_2(C)^i$. The last equation holds because $C$ is symmetric. Thus $\kappa_2(\hat{G}_i) \to \infty$. The rate of growth of the condition number is just enough to offset the convergence of the underlying subspace iterations.

In this example, the symmetry of the matrix $C$ is not essential. Every matrix of the form

$$\begin{bmatrix} a & c \\ b & 0 \end{bmatrix}, \qquad \text{where} \quad |a| < \min\{|b|, |c|\},$$

exhibits similar behavior: each step reverses the positions of the entries $b$ and $c$. These examples have analogues in higher dimensions. For example, the matrix

$$\begin{bmatrix} a & b & c \\ d & 0 & 0 \\ 0 & e & 0 \end{bmatrix}$$

exhibits cycling of the entries $c$, $d$, and $e$, provided that $a$ and $b$ are sufficiently small relative to $c$, $d$, and $e$. Matrices of this type do occasionally crop up in practice, and any practical implementation of the $LR$ algorithm with partial pivoting must have a mechanism for dealing with them.

For the other algorithms, which do not use any pivoting, the following type of strategy, called an *exceptional shift strategy*, is often employed. Each $p_i$ is chosen with an eye to making the subspaces converge. As the $i$th step is taken, the condition number of $G_i$ is monitored somehow. For example, the size of certain multipliers can be checked. If $G_i$ is found to be too ill conditioned, the step is restarted with a different choice of $p_i$. This is called an *exceptional* step. The hope is that if the $G_i$ are controlled, the $\hat{G}_i$ will not get too bad. Of course this is only a heuristic strategy, but it has been used with some success [9, 13]. The typical experience is that the exceptional steps are needed only in the early stages. Once the algorithm begins to converge, exceptional shifts are unnecessary.

It is of historical interest to mention one case (other than the $QR$ algorithm) for which the condition numbers of the transforming matrices can be guaranteed to remain bounded. Consider the stationary case ($p_i = p$) of the $LR$ algorithm without pivoting. Suppose $p(A_0)$ has eigenvalues of distinct modulus. Then if $A_0$ satisfies two other mild technical conditions, it can be shown that the sequence $(\hat{G}_i)$, which we denote $(\hat{L}_i)$ in this case, has a limit $\hat{L}$. Thus the condition numbers $\kappa_2(\hat{L}_i)$ are certainly bounded. Since the condition that the eigenvalues have distinct modulus (together with certain mild technical conditions) also guarantees that the subspaces con-

verge for $k = 1, \ldots, n-1$, the sequence $(A_k)$ of $LR$ iterates must converge to an upper triangular matrix. Rutishauser's original proof [for the case $p(A) = A$] of the convergence of the $LR$ algorithm [22] followed roughly these lines, although he used very different terminology. He showed that the $\hat{L}_i$ have a limit, then used that to conclude that the $A_i$ converge. Instead of using arguments involving subspaces, he used determinants. (The mild technical conditions that we have mentioned are subspace conditions like $\mathcal{S} \cap \mathcal{U} = \{0\}$. Rutishauser formulated them as conditions on determinants.)

*Choice of Shifts*

We now address the question of how to choose the $p_i$. There are many possible strategies. We will focus on the obvious strategy, one that usually works well in practice. After $i-1$ steps, we choose $p_i$ to be the characteristic polynomial of $A_{22}^{(i-1)}$, the trailing $m \times m$ submatrix of $A_{i-1}$. We will call this the *generalized Rayleigh-quotient shift strategy*, because in the case $m = 1$ it is just the Rayleigh-quotient shift. If $\|A_{21}^{(i-1)}\|_2$ is sufficiently small, the eigenvalues of $A_{22}^{(i-1)}$ will be good approximations to eigenvalues of $A_{i-1}$, so we expect this strategy to have good local convergence properties. Experience has shown that the global convergence is usually satisfactory as well, although no global convergence theorem is possible. There is a famous example [27, p. 362] for which the $QR$ algorithm with the generalized Rayleigh-quotient strategy fails to converge, regardless of the choice of $m$, excluding the ridiculous choice $m = n$. We also mention once again the chaotic behavior demonstrated by Batterson and Smillie [2] in the case $m = 1$. As for the local convergence, it is typically quadratic, as the following theorem shows.

THEOREM 6.3. *Let $A_0 \in \mathbb{C}^{n \times n}$ have distinct eigenvalues. Let $(A_i)$ be the sequence generated by the GR algorithm starting from $A_0$, using the generalized Rayleigh-quotient shift strategy with polynomials of degree $m$. Suppose there is a $\hat{\kappa}$ such that $\kappa_2(\hat{G}_i) \leq \hat{\kappa}$ for all $i$, and the $A_i$ converge to block triangular form, in the sense described in Theorem 6.2, with $k = n - m$. Then the convergence is quadratic.*

*Proof.* Let $\{\{\lambda_1, \ldots, \lambda_k\}, \{\lambda_{k+1}, \ldots, \lambda_n\}\}$ be any partition of the spectrum of $A_0$ into two subsets containing $k$ and $m$ elements, respectively, and let $\mathcal{T}$ and $\mathcal{U}$ be the invariant subspaces of $A_0$ associated with $\lambda_1, \ldots, \lambda_k$ and $\lambda_{k+1}, \ldots, \lambda_n$, respectively. Let $\mathcal{S}_0 = \langle e_1, \ldots, e_k \rangle$ and $\mathcal{S}_i = p_i(A_0)\mathcal{S}_{i-1}$, $i = 1, 2, 3, \ldots$, as in Theorem 6.2. We will show that there is a constant $M$, which depends only on $A_0$, such that for all sufficiently small $\epsilon > 0$, if $d(\mathcal{S}_i, \mathcal{T}) = \epsilon$, then $d(\mathcal{S}_{i+1}, \mathcal{T}) \leq M\epsilon^2$. This suffices to establish quadratic convergence.

Let $d(\mathcal{S}_i, \mathcal{T}) = \epsilon$. Applying Lemma 6.1 as in the proof of Theorem 6.2, we see that

$$\|A_{21}^{(i)}\|_2 \leq 2\sqrt{2}\,\hat{\kappa}\|A_0\|_2 d(\mathcal{S}_i, \mathcal{T}) = M_1\epsilon,$$

where $M_1 = 2\sqrt{2}\,\hat{\kappa}\|A_0\|_2$. The matrix $A_0$ is simple, so it has the form $A_0 = VDV^{-1}$, where $D = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$. Thus $A_i = V_i D V_i^{-1}$, where $V_i = \hat{G}_i^{-1}V$. The block triangular matrix

$$\begin{bmatrix} A_{11}^{(i)} & A_{12}^{(i)} \\ 0 & A_{22}^{(i)} \end{bmatrix} \tag{21}$$

is a perturbation of $A_i$ that differs from it by at most $M_1\epsilon$, so by the Bauer-Fike theorem [17], its eigenvalues differ from eigenvalues of $A_i$ by not more than $\kappa_2(V_i)M_1\epsilon \leq M_2\epsilon$, where $M_2 = \hat{\kappa}\kappa_2(V)M_1$. The polynomial $p_{i+1}$ to be used for the $(i+1)$st $GR$ step is

$$p_{i+1}(\lambda) = \prod_{l=1}^{m} \left(\lambda - \sigma_l^{(i+1)}\right),$$

where $\sigma_1^{(i+1)}, \ldots, \sigma_m^{(i+1)}$ are the eigenvalues of $A_{22}^{(i)}$. Since these are eigenvalues of (21), they are within $M_2\epsilon$ of $m$ eigenvalues of $A_i$, say

$$|\lambda_{k+l} - \sigma_l^{(i+1)}| \leq M_2\epsilon, \qquad l = 1, \ldots, m.$$

(By taking $\epsilon$ sufficiently small we can guarantee that no two of the $\sigma_l^{(i)}$ are within $M_2\epsilon$ of the same eigenvalue of $A_i$). If we make sure that $M_2\epsilon \leq 1$, then for $j = k+1, \ldots, n$,

$$\left|p_{i+1}(\lambda_j)\right| = \prod_{l=1}^{m} |\lambda_j - \sigma_l^{(i+1)}| \leq M_3\epsilon,$$

where $M_3 = (2\|A_0\|_2 + 1)^{m-1}M_2$. Let $\gamma = d(\{\lambda_1, \ldots, \lambda_k\}, \{\lambda_{k+1}, \ldots, \lambda_n\}) > 0$. As long as $M_2\epsilon \leq \gamma/2$, we have, for $j = 1, \ldots, k$, $|p_{i+1}(\lambda_j)| \geq (\gamma/2)^m$. Thus

$$\frac{\max\limits_{k+1 \leq j \leq n} |p_{i+1}(\lambda_j)|}{\min\limits_{1 \leq j \leq k} |p_{i+1}(\lambda_j)|} \leq M_4\epsilon, \tag{22}$$

where $M_4 = M_3(2/\gamma)^m$.

let $\tilde{\mathcal{S}}_i = V^{-1}\mathcal{S}_i$, $\tilde{\mathcal{T}} = V^{-1}\mathcal{T} = \langle e_1, \ldots, e_k \rangle$, and $\tilde{\mathcal{U}} = V^{-1}\mathcal{U} = \langle e_{k+1}, \ldots, e_n \rangle$. Then $\tilde{\mathcal{S}}_{i+1} = p_{i+1}(D)\tilde{\mathcal{S}}_i$. Define $T_1 \in \mathbb{C}^{k\times k}$ and $T_2 \in \mathbb{C}^{m\times m}$

by $T_1 = \operatorname{diag}\{\lambda_1, \ldots, \lambda_k\}$ and $T_2 = \operatorname{diag}\{\lambda_{k+1}, \ldots, \lambda_n\}$, so that $D = \operatorname{diag}\{T_1, T_2\}$. Then $p_{i+1}(T_1)$ is nonsingular by (22). By Lemma 4.2, $d(\tilde{\mathcal{S}}_i, \tilde{\mathcal{T}}) \leqslant \kappa_2(V)\epsilon$, so we can make $d(\tilde{\mathcal{S}}_i, \tilde{\mathcal{T}}) \leqslant \sqrt{3}/2$ by taking $\epsilon$ sufficiently small. This implies $\tilde{\mathcal{S}}_i \cap \tilde{\mathcal{U}} = \{0\}$ by Lemma 4.3. Thus we can apply Lemma 4.4 to obtain the inequality

$$d(\tilde{\mathcal{S}}_{i+1}, \tilde{\mathcal{T}}) \leqslant \frac{\beta}{\sqrt{1-\beta^2}} \|p_{i+1}(T_2)\|_2 \|p_{i+1}(T_1)^{-1}\|_2, \tag{23}$$

where $\beta = d(\tilde{\mathcal{S}}_i, \tilde{\mathcal{T}}) \leqslant \kappa_2(V)\epsilon$. Since $\beta \leqslant \sqrt{3}/2$, we have $\sqrt{1-\beta^2} \geqslant \frac{1}{2}$. Furthermore $\|p_{i+1}(T_1)^{-1}\|_2 = (\min_{1 \leqslant j \leqslant k} |p_{i+1}(\lambda_j)|)^{-1}$ and $\|p_{i+1}(T_2)\|_2 = \max_{k+1 \leqslant j \leqslant n} |p_{i+1}(\lambda_j)|$. Applying these results to (23), and using (22) and Lemma 4.2, we conclude that

$$d(\mathcal{S}_{i+1}, \mathcal{T}) \leqslant 2\kappa_2^2(V) M_4 \epsilon^2 = M\epsilon^2,$$

where $M = 2\kappa_2^2(V)^2 M_4$. ∎

For certain classes of matrices possessing special structure, the generalized Rayleigh-quotient strategy yields cubic convergence. In order to prove this we will need the following lemma, a variation on the Bauer-Fike theorem.

LEMMA 6.4. *Let* $A = \operatorname{diag}\{A_1, A_2\} \in \mathbb{C}^{n \times n}$ *be a block diagonal matrix whose blocks are simple*: $C_1^{-1} A_1 C_1 = D_1 = \operatorname{diag}\{\mu_1, \ldots, \mu_k\}$ *and* $C_2^{-1} A_2 C_2 = D_2 = \operatorname{diag}\{\mu_{k+1}, \ldots, \mu_n\}$. *Let*

$$E = \begin{bmatrix} 0 & E_{12} \\ E_{21} & 0 \end{bmatrix},$$

*where* $E_{21} \in \mathbb{C}^{(n-k)\times k}$. *If* $\lambda$ *is an eigenvalue of* $A + E$, *then*

$$\min_{1 \leqslant j \leqslant k} |\mu_j - \lambda| \min_{k+1 \leqslant j \leqslant n} |\mu_j - \lambda| \leqslant \kappa_2(C_1)\kappa_2(C_2)\|E_{12}\|_2 \|E_{21}\|_2. \tag{24}$$

*Proof.* Let $x$ be an eigenvector of $A + E$ associated with the eigenvalue $\lambda$. Let $C = \operatorname{diag}\{C_1, C_2\}$ and $z = C^{-1}x$. Then $C^{-1}(A + E)Cz = \lambda z$; that is,

$$\begin{bmatrix} D_1 & C_1^{-1} E_{12} C_2 \\ C_2^{-1} E_{21} C_1 & D_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \lambda \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

where we have partitioned $z$ in the obvious way. This equation implies

$$(\lambda I - D_1)z_1 = C_1^{-1}E_{12}C_2z_2$$

and

$$(\lambda I - D_2)z_2 = C_2^{-1}E_{21}C_1z_1.$$

If either $\lambda I - D_2$ or $\lambda I - D_1$ is singular, (24) is trivially true. Otherwise

$$z_1 = (\lambda I - D_1)^{-1}C_1^{-1}E_{12}C_2(\lambda I - D_2)^{-1}C_2^{-1}E_{21}C_1z_1.$$

Taking norms of both sides and dividing by $\|z_1\|_2$, we obtain

$$1 \leqslant \|(\lambda I - D_1)^{-1}\|_2 \|(\lambda I - D_2)^{-1}\|_2 \kappa_2(C_1)\kappa_2(C_2)\|E_{12}\|_2\|E_{21}\|_2,$$

from which (24) follows. ∎

THEOREM 6.5. *Under the hypotheses of Theorem* 6.3, *suppose that each of the iterates*

$$A_i = \begin{bmatrix} A_{11}^{(i)} & A_{12}^{(i)} \\ A_{21}^{(i)} & A_{22}^{(i)} \end{bmatrix}$$

*satisfies* $\|A_{12}^{(i)}\| = \|A_{21}^{(i)}\|$ *for some fixed norm* $\|\cdot\|$. *Then the iterates converge cubically if they converge.*

*Proof.* The proof is the same as that of Theorem 6.3, except that in this case the shifts, which are eigenvalues of

$$\begin{bmatrix} A_{11}^{(i)} & 0 \\ 0 & A_{22}^{(i)} \end{bmatrix}, \tag{25}$$

can be shown to differ from eigenvalues of $A_i$ by only $O(\epsilon^2)$. First of all, by the Bauer-Fike theorem we can show, just as in the proof of Theorem 6.3, that there is a constant $M_2$, independent of $i$, such that the eigenvalues of (25) differ from those of $A_i$ by not more than $M_2\epsilon$. By taking $\epsilon$ sufficiently small we can guarantee that no two eigenvalues of (25) are within $M_2\epsilon$ of the

same eigenvalue of $A_i$. Let $\mu_1,\dots,\mu_k$ and $\mu_{k+1},\dots,\mu_n$ denote the eigenvalues of $A_{11}^{(i)}$ and $A_{22}^{(i)}$ respectively, and let $\lambda_1,\dots,\lambda_n$ denote the eigenvalues of $A_i$, ordered so that $|\lambda_j-\mu_j|\leqslant M_2\epsilon$ for $j=1,\dots,n$. Letting $\gamma = d(\{\lambda_1,\dots,\lambda_k\},\{\lambda_{k+1},\dots,\lambda_n\})$, assume that $\epsilon$ is small enough that $|\lambda_j-\mu_h|\geqslant \gamma/2$ for $h=1,\dots,k$ and $j=k+1,\dots,n$. Applying Lemma 6.4 with $A$ given by (25), $A+E=A_i$, and $\lambda=\lambda_j$, $k+1\leqslant j\leqslant m$, we find that

$$|\mu_j-\lambda_j| = \min_{k+1\leqslant h\leqslant n}|\mu_h-\lambda_j| \leqslant \frac{2}{\gamma}\kappa_2(C_1)\kappa_2(C_2)\|A_{21}^{(i)}\|_2\,\|A_{12}^{(i)}\|_2,$$

for $j=k+1,\dots,n$, where $C=\operatorname{diag}\{C_1,C_2\}$ is a matrix that diagonalizes $\operatorname{diag}\{A_{11}^{(i)},A_{22}^{(i)}\}$. Certainly such a $C$ exists, provided $\epsilon$ is sufficiently small. From Theorems 3 and 5 of [23] it follows that $\kappa_2(C_1)\kappa_2(C_2)$ can be bounded above independently of $i$, provided $\epsilon$ is sufficiently small (i.e. $i$ is sufficiently large). Since the norms $\|A_{21}^{(i)}\|_2$ and $\|A_{12}^{(i)}\|_2$ are both of order $\epsilon$, there must be a constant $M_5$ such that

$$|\mu_j-\lambda_j|\leqslant M_5\epsilon^2, \qquad j=k+1,\dots,n.$$

Since $\mu_{k+1},\dots,\mu_n$ are exactly the shifts $\sigma_1^{(i+1)},\dots,\sigma_m^{(i+1)}$, we can now proceed as in the proof of Theorem 6.3 to obtain

$$\frac{\max_{k+1\leqslant j\leqslant n}|p_{i+1}(\lambda_j)|}{\min_{1\leqslant j\leqslant k}|p_{i-1}(\lambda_j)|}\leqslant M_6\epsilon^2,$$

in analogy with (22), and ultimately $d(\mathscr{S}_{i+1},\mathscr{T})\leqslant M\epsilon^3$. ∎

REMARK. It is clear from the proof that we have not used the full strength of the assumption $\|A_{21}^{(i)}\| = \|A_{12}^{(i)}\|$. All that is really needed is that $\|A_{12}^{(i)}\|\leqslant M_7\|A_{21}^{(i)}\|$ for some $M_7$ independent of $i$.

EXAMPLE 6.6. If we apply the $QR$ algorithm to a normal matrix $A_0$, then all $A_i$ will be normal. Hence they will satisfy $\|A_{12}^{(i)}\|_F = \|A_{21}^{(i)}\|_F$. Thus the $QR$ algorithm with the generalized Rayleigh-quotient shift strategy applied to normal matrices converges cubically when it converges.

EXAMPLE 6.7. A matrix $A\in\mathbb{R}^{2n\times 2n}$ is called *Hamiltonian* if it satisfies $(JA)^T=JA$, where $J$ is as defined in Example 2.4. The $SR$ algorithm

preserves Hamiltonian matrices. It is this property that makes the $SR$ algorithm useful. The algebraic Riccati equation is a disguised eigenvalue problem whose matrix is Hamiltonian. One can preserve this important property by using the $SR$ algorithm [9, 10]. Hamiltonian matrices satisfy $\|A_{12}^{(i)}\|_F = \|A_{21}^{(i)}\|_F$, provided $m$ is even. This is not a serious restriction; the eigenvalues of Hamiltonian matrices always occur in pairs, so $m$ should always be taken to be even. Thus the convergence rate of the Hamiltonian $SR$ algorithm is typically cubic.

EXAMPLE 6.8. Let $\mathcal{J}$ be the subgroup of $GL_n(\mathbb{C})$ whose members are the diagonal matrices whose main diagonal entries lie in $\{1, -1\}$. Given $J \in \mathcal{J}$, a matrix $A \in \mathbb{C}^{n\times n}$ is said to be *J-Hermitian* [*J-skew-Hermitian*] if $(JA)^* = JA$ [$(JA)^* = -JA$]. Let $J_a, J_b \in \mathcal{J}$ have the same inertia. A matrix $H \in \mathbb{C}^{n\times n}$ is called $(J_a, J_b)$-*unitary* if $H^*J_aH = J_b$. If $A$ is $J_a$-Hermitian [$J_a$-skew-Hermitian] and $H$ is $(J_a, J_b)$-unitary, then $H^{-1}AH$ is $J_b$-Hermitian [$J_b$-skew-Hermitian]. The $HR$ algorithm produces a sequence $(A_i)$ by $A_i = H_i^{-1}A_{i-1}H_i$, where $H_i$ is $(J_{i-1}, J_i)$-unitary, for some $(J_i)$. We cannot control the $J_i$, except that we get to choose $J_0$. The accumulated transforming matrix $\hat{H}_i = H_1 \cdots H_i$ is $(J_0, J_i)$-unitary. If $\mathbf{A_0}$ is $J_0$-Hermitian [$J_0$-skew-Hermitian], then $A_i$ is $J_i$-Hermitian [$J_i$-skew-Hermitian]. Matrices with any of these symmetries satisfy $\|A_{12}^{(i)}\|_F = \|A_{21}^{(i)}\|_F$.

EXAMPLE 6.9. Complex symmetric and skew-symmetric matrices are preserved by the complex, orthogonal $QR$ algorithm of Cullum and Willoughby [12]. Matrices of either of these forms obviously satisfy $\|A_{12}^{(i)}\|_F = \|A_{21}^{(i)}\|_F$.

## REFERENCES

1 Z. Bai and J. Demmel, On a Block Implementation of Hessenberg Multishift $QR$ Iteration, LAPACK Working Note No. 8, Argonne National Lab. MCS-TM-127, Jan. 1989.

2 S. Batterson and J. Smillie, Rayleigh quotient iteration for nonsymmetric matrices, *Math. Comp.*, to appear.

3 F. L. Bauer, Das Verfahren der Treppeniteration und verwandte Verfahren zur Lösung algebraischer Eigenwertprobleme, *Z. Angew. Math. Phys.* 8:214–235 (1957).

4 F. L. Bauer, On modern matrix iteration processes of Bernoulli and Graeffe type, *J. Assoc. Comput. Mach.* 5:246–257 (1958).

5 Å. Björck and G. Golub, Numerical methods for computing angles between linear subspaces, *Math. Comp.* 27:579–594 (1973).

6 M. A. Brebner and J. Grad, Eigenvalues of $Ax = \lambda Bx$ for real symmetric matrices $A$ and $B$ computed by reduction to a pseudosymmetric form and the $HR$ process, *Linear Algebra Appl.* 43:99–118 (1982).
7 W. Bunse and A. Bunse-Gerstner, *Numerische Lineare Algebra*, Teubner, Stuttgart, 1985.
8 A. Bunse-Gerstner, An analysis of the $HR$ algorithm for computing the eigenvalues of a matrix, *Linear Algebra Appl.* 35:155–178 (1981).
9 A. Bunse-Gerstner and V. Mehrmann, A symplectic $QR$ like algorithm for the solution of the real algebraic Riccati equation, *IEEE Trans. Automat. Control* AC-31:1104–1113 (1986).
10 A. Bunse-Gerstner, V. Mehrmann, and D. Watkins, An $SR$ algorithm for Hamiltonian matrices based on Gaussian elimination, *Methods Oper. Res.* 58:339–358 (1989).
11 H. J. Buurema, A Geometric Proof of Convergence for the $QR$ Method, Doctoral Dissertation, Univ. of Groningen, 1970.
12 J. Cullum and R. Willoughby, A $QL$ algorithm for complex, symmetric tridiagonal matrices, preprint, IBM Thomas J. Watson Research Center, 1987.
13 A. Dax and S. Kaniel, The $ELR$ method for computing the eigenvalues of a general matrix, *SIAM J. Numer. Anal.* 18:597–605 (1981).
14 J. Della-Dora, Numerical linear algorithms and group theory, *Linear Algebra Appl.* 10:267–283 (1975).
15 J. J. Dongarra et al., *Linpack User's Guide*, SIAM, Philadelphia, 1979.
16 L. Elsner, On some algebraic problems in connection with general eigenvalue algorithms, *Linear Algebra Appl.* 26:123–138 (1979).
17 G. Golub and C. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins, U.P., Baltimore, 1989.
18 T. Kato, *Perturbation Theory for Linear Operators*, corrected printing of 2nd ed., Springer-Verlag, New York, 1980.
19 B. N. Parlett, Global convergence of the basic $QR$ algorithm on Hessenberg matrices, *Math. Comp.* 22:803–817 (1968).
20 B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
21 B. N. Parlett and W. G. Poole, Jr., A geometric theory for the $QR$, $LU$, and power iterations, *SIAM J. Numer. Anal.* 8:389–412 (1973).
22 H. Rutishauser, Solution of eigenvalue problems with the $LR$-transformation, *Nat. Bur. Standards Appl. Math. Ser.* 49:47–81 (1958).
23 R. A. Smith, The condition numbers of the matrix eigenvalue problem, *Numer. Math.* 10:232–240 (1967).
24 D. S. Watkins, Understanding the $QR$ algorithm, *SIAM Rev.* 24:427–440 (1982).
25 D. S. Watkins and L. Elsner, Chasing algorithms for the eigenvalue problem, *SIAM J. Matrix Anal. Appl.*, to appear.
26 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford U.P., 1965.
27 J. H. Wilkinson and C. Reinsch, *Handbook for Automatic Computation, Vol. II, Linear Algebra*, Springer-Verlag, New York, 1971.

*Received 2 March 1989; final manuscript accepted 1 December 1989*